Peer Review File

Manuscript Title: Multi-qubit gates and Schrödinger cat states in an optical clock

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

In “Multi-qubit gates and ‘Schrodinger cat’ states in an optical clock”, Cao et al report on GHZ states created on the optical clock transition in Strontium atoms, arranged by tweezers into programmable configurations within a two-dimensional lattice potential. New global Rydberg pulse techniques are developed using optimal control methods in order to create large, programmable clusters of GHZ states and the GHZ states are characterized and their fidelities compared to some of the known sources of imperfections. By parking at a short dead time, GHZ states are shown to have lower Allan deviation than the standard quantum limit, and some comparisons to the Heisenberg scaling limit are analyzed. Finally, by using cascaded arrays of GHZ states of various sizes (including postprocessed grouping of data), a cascaded phase tracking protocol is implemented to extend the dynamic range of phase detection, as tested by deterministic injection of specific phases onto the laser. As the first implementation and characterization of some of these central protocols for entanglement-enhanced metrology, I recommend publication in Nature. However, before accepting I have several questions / suggestions I would like the authors to address.

1. My main technical comments regard the authors’ new GHZ state creation protocol and its characterization. It is of course evident in Fig 2c that the GHZ fidelity is significantly worse than the plotted contributions to decoherence. While the data is presented honestly, for this new protocol I ask that the authors provide more analysis or estimates as to what is limiting this protocol, as others will attempt to use these techniques to create GHZ states for various purposes. I think such analysis would strengthen the authors’ manuscript and its impact on the tweezer community.
a. The authors correctly identify that blockade violation effects will be a dominant issue for this protocol. However, more analysis and commentary is prudent. In ED Table I, the listed blockade strengths are actually quite large – the minimum interaction strength being 9 Omega. Are the authors’ protocols being limited by blockade violations (which would be apparent experimentally for example as anomalously large presence in blockade-violating states), or more so the fact that their pulse is calculated for perfect blockade, and e.g. the effective light shifts of the form Omega^2 / V are degrading the pulse fidelity? How is the pulse calibrated, is it just blindly implemented based on the numerically found pulse? To put in other words, are the gray points in Fig 2c coming only from the calculated blockade violation (population left in Rydberg), or are they a simulation that includes the actual Rydberg interaction strengths between all pairs? If it’s in fact a full simulation with the actual pair interaction strengths, then the authors should more explicitly state this. The authors should also clarify that, if these Omega^2/V light shifts are a limiting factor, that these can be partially compensated by designing a pulse that is calculated at the finite interaction strengths used. More discussion and explanation of these various points would be useful.

b. What is the “spaghetti distance” under the authors’ present configuration? I.e., the distance where if two atoms are too close, then the blockade can again begin to be compromised? (due to the fact that pair interaction strengths start to become comparable to the splitting between pair states). I would speculate the authors are quite close to this as they put the atoms exceptionally close (down to only 1 lattice spacing?). This should also be mentioned and discussed.
c. The “spaghetti distance”, setting how close two atoms can be placed, and the finite blockade, will set a fundamental limit on how large of GHZ states of a certain fidelity can be produced with the authors’ techniques (also accounting for the enhanced rabi frequency sqrt(N) becoming increasingly comparable to the finite interaction strength). This should be estimated and discussed, as for these reasons it is fundamentally not clear if the authors’ methods will scale to much larger GHZ states.
d. Regardless, it does not seem like the GHZ infidelity is primarily arising from finite blockade, as gauged by Figure 4c if I correctly understand the configuration (although it depends on the specific atom distances). In particular, is it not interesting that here the GHZ state fidelity is only slightly decreasing with increasing N? E.g., in this figure with the same total pulse length, the 4 qubit GHZ fidelity is much better than (2 qubit GHZ fidelity)^2. What does this imply about the underlying error mechanisms that are limiting the fidelity?
e. Relatedly, as there are clearly other decoherence sources, there should be some estimation of these and how they affect this protocol. It is OK if these are rough estimates and not detailed atomic physics calculations, but e.g. is Doppler a major contribution here? What about the positional spread from the drop-recapture (with a combo of concerns on “spaghetti” and finite interaction strengths)? Practitioners and theorists will consider the authors’ GHZ protocol for various applications with atom arrays, and such discussion / estimates would increase the impact of their manuscript.
f. The authors state “The calculated gate duration TG scales sublinearly in Nmax, potentially enabling improved GHZ-state fidelities compared to successive application of two-qubit gates when gate errors are dominant.” This is not clear, and depends on the error mechanisms (although I appreciate the use of the word “potentially”). An N-qubit GHZ state creation circuit requires N (parallelizable) gates, but log(N) depth, compared to the N^0.59 scaling found here. So , e.g., if fidelity decays with total pulse time (as is seemingly being observed here) and not e.g. total time spent in the Rydberg state, then wouldn’t the log depth GHZ circuit be better? Some discussion along these lines would be useful.
g. To summarize all the above, the authors manuscript would benefit from more detailed description of the limitations of this approach to creating GHZ states. Although the authors’ protocol benefits from simplicity of global pulses, there are (a) fundamental limits on size (which the authors are approaching) given by the tradeoff between “spaghetti distance” and “Blockade range”, (b) there is (as stated) a significant contrast reduction for smaller ensembles, and (c) for error mechanisms which scale with pulse length (as appears to be happening here) then naively a log(N) circuit would be better than the N^0.59 pulses found here. In contrast, the authors produce a very high fidelity Bell state, likely corresponding to a gate fidelity > 0.99, which in theory should allow them to realize GHZ states of ~50 atoms. The bottom-line: when analyzing these various considerations – is it better or worse to make GHZ states in this way as compared to a conventional log depth circuit? The authors’ results and new pulse techniques still stand regardless, but clearly reporting on these aspects would be very beneficial to the readership.

2. Related to the above, the authors should provide a brief comparison, at least qualitative, to spin squeezing for metrology (including their prior work). It would be interesting to have even a

qualitative perspective on, given the results in the present manuscript, does it appear squeezing or cascaded GHZ states will be more effective for sensing in the near (and long) term?

3. I ask the authors include their new pulses as some .csv or .txt supplementary files with the published version of the manuscript, if this is not already the case.

4. Throughout the manuscript the authors say they are doing X(pi/2) gates. I believe the authors mean that they are doing Y(pi/2) gates, as a rotation around the Y axis is what creates a superposition |0> + |1>, whereas a rotation around the X axis creates |0> + i |1>. Although just notation, please check this and ensure consistency with the literature.

5. Technical question: the authors describe their intensity modulation when scanning phase of the AOM. Why do the authors not just double-pass their AOM? In order to preserve optical power? Furthermore, are the 5-10% intensity modulations not compensated at all? Would this not give a massive miscalibration error? How are the pulses being calibrated, and can miscalibration simply be the source of the fidelity degradation with GHZ size? Please comment on this.

6. Can the authors describe how much power is on the atoms and what the size of the Rydberg beams are, for their 4 MHz rabi frequencies?

7. With respect to metrology, this work provides very valuable proof-of-principle demonstrations. However, I think the presentation can be improved and it can be clarified in multiple places that these results are proof-of-principle.

a. The authors present results beyond SQL. However, the GHZ state coherence times are very short, and are not even limited by laser phase noise but by unrelated global correlated magnetic field noise. As such the authors measure at a short dark time of 3 ms. Although the GHZ outperforms the CSS for this particular dark time, can the authors justify that this is truly “below-SQL performance”? Since, of course, one should measure with a dark time ~ coherence time, and the GHZ state has quite reduced coherence times. My concern about this point can be fixed by appropriate wording modifications in the manuscript.
b. The inset for Fig 3b is confusing. It initially appears that the authors are saying the CSS and SQL are scaling with the same slope as the Heisenberg limit. I suggest making the presentation clearer. Relatedly, as plotted it appears that the grey circles are approaching the HL scaling, and not the other way around that the HL scaling is modified to be the gray circles.
c. The statement “These gains can be practically harnessed for a restricted class of problems…”, is this still true here given that the reduction in coherence time is not even coming from laser phase noise, but from unrelated global magnetic field noise?
d. If I understand correctly, the authors operate with up to 30 atoms at a time, but then postprocess data to include multiple runs as “simultaneous ensembles” and get to a “total atom number” of 118 atoms. Since these are proof-of-principle experiments, and this bootstrapping allows for probing physics of the protocol, I think it is ok. However, “total atom number” could mislead readers and should be re-worded to make it clear that this is e.g. “postprocessed atom number”. Technically this also means that previous experiments where people made variable-sized GHZ states but in different shots could also be postprocessed to do this cascaded GHZ state protocol. I think it’s fine, but the

authors should word more carefully.
e. The theoretical dynamic range and mean-square error of Nk = 8 are plotted in Fig 4d and 4e. However, don’t the authors have the data to just experimentally plot these? Is there a reason that they did not? This is not necessary but would be interesting to see experimental data here.
f. Is 4f assuming something particular about the laser phase noise distribution? If so, please state it.

Referee #2 (Remarks to the Author):

Summary
In their manuscript “Multi-qubit gates and ‘Schrödinger cat’ states in an optical clock” the authors demonstrate their advance towards realistic metrological gain of using entanglement in atomic optical clocks with tweezer arrays. The authors provide several suites of characterization of their efforts, including the generation of GHZ-states up to size 9, a cascade of GHZ-states of increasing size, and detailed characterization of the phase estimation properties of both of these. The authors supplement their main text with a comprehensive methods section that enables a deeper understanding of the material.

High-level feedback
My overall assessment of the manuscript in its current state is that it constitutes an impressive effort, written in an accessible and at the same time accurate fashion. The figures are both informative and beautiful, though at times I feel like in print a magnifying glass will be required. The authors have successfully navigated many of the pitfalls about statements in quantum metrology, which I greatly appreciate. The manuscript has clearly been diligently proof-read; I can find only a single potential typo. This work has been made possible by the application of technical and theoretical advances of recent years and is now pushing the field of quantum metrology further. I do have a number of comments that I believe will improve the clarity of the manuscript, but if it were published as is today it would still constitute one of the better Nature papers I have ever read. It was a pleasure reading it and I congratulate everyone involved.

Detailed feedback

Abstract:
There are two things that come to mind in the abstract. Firstly “sufficiently short dark times” is perhaps the main issue with quantum-enhanced phase estimation. You win linearly in Ramsey time while you win *at best* linearly with number of probes, so in general you lose when using entanglement if you account for increase in duty cycle etc. . While the authors address this in the manuscript to an extent, it is an important point that ought to be highlighted – perhaps in the discussion section rather than the abstract. The second thing is the jump from “perform below SQL” for up to 4 qubits and then moving on to “A key challenge”. This is also something that I struggled with in the manuscript main text and I will get to that at the appropriate section. However, I think it is not sufficiently clear how the two sentences immediately following each other are related. One might wonder why below SQL is not enough, for example. See comments to Fig. 4.

Introduction:
I had expected to see a reference to the Oxford paper on entangled clocks in the introduction. This is something the authors may consider adding.
l. 15: Here the Heisenberg limit is equated with the 1/N scaling and it is explicitly stated that this is the fundamental bound given by quantum theory. If that is so, then we cannot explain "super-Heisenberg" scaling within quantum theory, which is of course not the case. Instead, Heisenberg scaling for linear observables is limited to 1/N and what people call "super-Heisenberg" is just Heisenberg for a nonlinear observable. I would advise adding the single word "linear" to stay correct outside of the most-common case of linear observable (here \sigma_z) and stand against the "super-Heisenberg" misnomer.
l. 24: One could expand the list with use-cases of real-life application from the somewhat arcane gravity and dark matter to life science with "Quantum-enhanced nonlinear microscopy" also in Nature, in particular in view of the wider readership of Nature.
l. 43: I am unsure what exactly 'scalable' means here. In particular I am not sure if it is true if I just use my own interpretation without context due to the distance scaling of the Rydberg blockade. Or is this meant more generally that in principle entanglement can assist at this front? Overall, I am not sure where this sentence is going.
l. 47: I am not sure what "macroscopically distinct" means in this context. I feel like you mean "mutually orthogonal"? As in 'x' or 'not x'. Nothing here is macroscopic?
l. 50: I was surprised by this "thus". Is the fact that the GHZ-state has N-times faster oscillation in parity actually the reason for saturating the HL? I thought both facts are true, but not that one causes the other.
l. 67: Here the authors mention that this is the first time that GHZ states are used for below-SQL performance. I am wondering whether they are being a very 'technically correct' or missing references. Both [3] and [63] could argue that they have done so. [3] has at admittedly extremely short Ramsey times and self-referenced which one may reasonably have reservations against. [63] has shown at least parts of the performance below the SQL. Of course, neither have done GHZ but instead squeezed states or optimal probe states, but the GHZ is just a squeezed state for a particular squeezing parameter.
Fig. 1: An indication that it is an optical qubit would be good, in particular since some other publications insist on stressing this. In the caption either I cannot parse the "form from" or it is a typo. You mean the "shape from" or is it "from" twice? The feature sizes from the fluorescence image are going to challenge any printer and researcher above the age of 25 in print. As for the circuit: Either you use some definition I have never seen or this circuit has one typo. This thing makes |01> + |10> when I do it by hand, matrix multiplication or in qiskit. If you start in |11> it would be the GHZ or if you have either of the X-rotations with a negative sign. Matters not for what you want to show, of course, but you write explicitly it is |00> + |11>.
l. 76: The authors use a Ramsey time of 3 ms, which is substantially below their coherence time. As discussed above, Ramsey time wins linearly in fractional instability so should always be as long as possible. I do not know what the rule of thumb is for how much below your coherence time you should stay before the Ramsey time starts to hurt more than aid. Is the dark time chosen such that at N = 4 you are just below this rule? If so, state that? If not, why are you at 1% of your coherence time?
GHZ design and GHZ-state preparation:
l. 103: This entire sentence was very confusing, because n is used inconsistently. First "\sqrt{n}-

enhanced”, so n is clearly a number. Then n is the sum of projectors, so an operator. Can one of them get a hat? Overall, I was repeatedly confused because I forgot what n and N are. Perhaps the operators can get a different symbol or some such.
l. 107: I was perplexed for a while as to what the Rabi trajectories are. When I hear trajectories in the context of gates I think of phase space trajectories, which is probably not what is meant here. Instead, do the authors mean that the different sub-ensembles see different laser parameters because they are spatially separated and therefore they use a pulse that is insensitive to fluctuations of these parameters? That is two ensembles at different locations see, say, different Rabi rate so you chose a pulse that is robust against Rabi rate changes to implement your 2\pi pulse up and down to |r>?
l. 136: Again, I think that makes the wrong Bell state unless it is -\pi/4 or you start in |1> or it is -\pi/2 in the start, no?
l. 172 / Fig. 2: I was perplexed by this graph until I realized grey points are not grey markers. Perhaps call out the dashed line or give one of those a different colour.

GHZ-state atom-laser comparison:
l. 226: Here again the question arises why it is 3 ms and not more, as discussed above.
Fig. 3: The fluorescence images are very small. It took me a fair bit to parse the inset. I concluded that this is “how much better am I than”, but perhaps this could be said more directly. It is all there, I know, but I still had trouble with it.

Cascaded GHZ-state phase estimation:
l. 347: This section makes no mention of what T is used. Is it still 3 ms? If so, that should be stated somewhere in the section or figures.
Fig. 4: Looking at sub-figure f the question arises how both abstract and the preceding section claim that the authors have beat the SQL but then here and in the main text of this section they say that the scheme is still not at the SQL by a 2 dB margin. Is the statement that with the short Ramsey times above you beat the SQL, but in a realistic scenario you would not with a single GHZ be able to due to the limited dynamic range? If so, then this section would be about using a larger Ramsey time, but that is not stated. If not and you are still at 3 ms then this section would be an “in principle we would need to” and then the limit would be something else. For example, because it is harder to make several ensembles rather than one. Is this the case? Either way, that bit was unclear to me from the very start and the presentation should be sharpened to chisel out this point more clearly.

Conclusion:
l. 387: I just wanted to point out that VQA on hardware is typically quite demanding in terms of number of shots, which for cold atoms are the costliest of all platforms. If the many iterations typically required on the platform are not run substantially faster than drift rates or if the fluctuations during the optimization are sufficiently large then this is going to be a fruitless endeavour. Which is not to say that it cannot be done or the authors should not mention it, only that it is very hard to do well.
I would have liked a short discussion on the topic of Ramsey times in the seconds for optical clocks with uncorrelated particles versus entanglement-enhanced ones given that the coherence time falls quickly with ensemble size.

Methods:

l. 1015: The sequence here is correct but it is not the one in the figure.

l. 1022: That is the one in the figure and here the authors correctly say that it produces the (other) Bell state. Generally, I find calling one of them “the Bell state” confusing, since there are four and the GHZ is one of them.

Eq. 8 / l. 1055: You write |+_x> in the equation but |+_y> in the text. As far as I can see, the text is correct, otherwise you are missing a final z-rotation.

l. 1063: You do not mean all of U is just a number, correct? You mean the prefactor before the operator exponential, no?

l. 1107: The authors say that the rearrangement success rate varies between 85 – 98%. Do I know before measurement? What does this mean for the duty cycle? I believe the authors allude several times to there being fewer than expected atoms in some GHZ states, which would mean you do not know beforehand and just live with it being worse than anticipated?

Author Rebuttals to Initial Comments:

Reply to referees for manuscript “Multi-qubit gates and ‘Schrodinger cat’ states in an optical clock”

June 12, 2024

Response to referee 1

> In "Multi-qubit gates and ‘Schrodinger cat’ states in an optical clock", Cao et al report on GHZ states created on the optical clock transition in Strontium atoms, arranged by tweezers into programmable configurations within a two-dimensional lattice potential. New global Rydberg pulse techniques are developed using optimal control methods in order to create large, programmable clusters of GHZ states and the GHZ states are characterized and their fidelities compared to some of the known sources of imperfections. By parking at a short dead time, GHZ states are shown to have lower Allan deviation than the standard quantum limit, and some comparisons to the Heisenberg scaling limit are analyzed. Finally, by using cascaded arrays of GHZ states of various sizes (including postprocessed grouping of data), a cascaded phase tracking protocol is implemented to extend the dynamic range of phase detection, as tested by deterministic injection of specific phases onto the laser. As the first implementation and characterization of some of these central protocols for entanglement-enhanced metrology, I recommend publication in Nature. However, before accepting I have several questions / suggestions I would like the authors to address.

We thank the referee for their positive characterization of the manuscript and their detailed feedback.

> 1. My main technical comments regard the authors’ new GHZ state creation protocol and its characterization. It is of course evident in Fig 2c that the GHZ fidelity is significantly worse than the plotted contributions to decoherence. While the data is presented honestly, for this new protocol I ask that the authors provide more analysis or estimates as to what is limiting this protocol, as others will attempt to use these techniques to create GHZ states for various purposes. I think such analysis would strengthen the authors’ manuscript and its impact on the tweezer community.

We thank the referee for this and the following comments regarding errors for the multi-qubit gate GHZ-state preparation. Indeed these are crucial considerations for the utility of this technique. As we discuss in further detail below, we have added an additional section dedicated to this in the Methods called “Sources of error in GHZ-state preparation”, starting on line 1203; this includes explicit error modeling of commonly known errors for the Bell state and 4-atom GHZ state, alongside a discussion of improvements that can be made to mitigate other error sources which are less well controlled and characterized. This section is accompanied by a new Extended Data Figure 3 showing the sensitivity of the GHZ state fidelity to certain errors as well as the results of the error modeling.

> a. The authors correctly identify that blockade violation effects will be a dominant issue for this protocol. However, more analysis and commentary is prudent. In ED Table I, the listed blockade strengths are actually quite large - the minimum interaction strength being 9

> Omega. Are the authors' protocols being limited by blockade violations (which would be apparent experimentally for example as anomalously large presence in blockade-violating states), or more so the fact that their pulse is calculated for perfect blockade, and e.g. the effective light shifts of the form Ω^2/V are degrading the pulse fidelity? How is the pulse calibrated, is it just blindly implemented based on the numerically found pulse? To put in other words, are the gray points in Fig 2c coming only from the calculated blockade violation (population left in Rydberg), or are they a simulation that includes the actual Rydberg interaction strengths between all pairs? If it's in fact a full simulation with the actual pair interaction strengths, then the authors should more explicitly state this. The authors should also clarify that, if these Ω^2/V light shifts are a limiting factor, that these can be partially compensated by designing a pulse that is calculated at the finite interaction strengths used. More discussion and explanation of these various points would be useful.

The gray points shown in Fig. 2c are coming from a full ED simulation taking into account the varying interaction between different pairs in the ensemble (i.e. $V_{ij} = C_6/r_{ij}^6$ for different spatial separations r_{ij} between atoms i and j). Closer inspection of the simulation result shows a largely negligible excitation ($\leq 10^{-4}$) of multiply excited Rydberg states up to the maximum $N = 9$ size investigated. This implies that the infidelity largely arises from effects such as the effective Ω^2/V light-shift as the referee suggests. We have revised this discussion in the main text to be more clear and reflect these nuances (lines 170–181).

Our pulses are currently calculated ignoring the role of the finite blockade. While we agree that one can try to optimize the pulse for a given blockade parameter, it is not obvious that such a strategy will perform well for these multi-qubit gates. Including the varying blockade will make the numerical pulse optimization significantly more expensive (increasingly so for larger ensembles), and even then the global laser pulse fundamentally cannot fully compensate for dynamics out of the symmetric subspace [1]. As outlined in Ref. [1], an alternative strategy is to optimize for robustness relative to the infinite blockade limit as opposed to for a specific blockade parameter; the pulses produced by this approach largely rely on reducing the Rabi frequency, however, a point already mentioned in the text.

> b. What is the "spaghetti distance" under the authors' present configuration? I.e., the distance where if two atoms are too close, then the blockade can again begin to be compromised? (due to the fact that pair interaction strengths start to become comparable to the splitting between pair states). I would speculate the authors are quite close to this as they put the atoms exceptionally close (down to only 1 lattice spacing?). This should also be mentioned and discussed.

As a proxy for the "spaghetti distance," we estimate the largest spacing at which a two-Rydberg-atom molecular state comes into resonance with twice the energy of the un-blockaded Rydberg state, as studied in Ref. [2]. We refer to this distance as $R_\times$. For the $47\,^3S_1$ Rydberg state used in this work, we calculate that $R_\times \approx 885\,\mathrm{nm}$ [3]. This means that atoms separated by two or more lattice spacings (as was always the case for the $N = 2, 4$ ensembles used) should be outside of the "spaghetti". On the other hand, atoms separated by one lattice spacing (present for $N = 6, 8, 9$ ensembles) are in a regime where molecular resonances could be significant, but the distances at which one expects these resonances are challenging to accurately calculate. Empirically, we have not observed signatures of any molecular resonances when performing spectroscopy on the blockaded Rydberg transition with two atoms separated by a single lattice site, although we cannot rule out weak off resonant effects leading to commensurately small errors. An exploration of these resonances would be an interesting avenue for future investigation. We have added a statement in the new Methods section 'Sources of error in GHZ-state preparation' with a citation to Ref. [2] to point out the potential role of complicated molecular spectrum at small inter-nuclear separations (lines 1244–1248).

> c. The "spaghetti distance", setting how close two atoms can be placed, and the finite blockade, will set a fundamental limit on how large of GHZ states of a certain fidelity can be produced with the authors' techniques (also accounting for the enhanced rabi frequency $\sqrt(N)$ becoming increasingly comparable to the finite interaction strength). This should be estimated and discussed, as for these reasons it is fundamentally not clear if the authors' methods will scale

> to much larger GHZ states.

As the referee points out, there may be a minimum usable atomic spacing $R_{\min}$ – potentially set by the "spaghetti distance" – as well as a maximum spacing $R_{\max}$ – set by the Rydberg blockade strength. These constraints set a maximum atom number N_b, which limits the maximum GHZ-state size. For the purposes of a rough scaling argument including potential spaghetti effects, we can assume that $R_{\min} \propto R_{\times} \propto n^{8/3}$ [2] (only in these paragraphs we use n to refer to principal quantum number). Taking $R_{\max}$ to scale as the blockade radius and given the scalings $C_6 \propto n^{11}$ and $\Omega_r \propto n^{-3/2}$ [4], this gives $R_{\max} \propto (C_6/\Omega_r)^{1/6} \propto n^{25/12}$. In two dimensions, one finds that $N_b \propto (R_{\max}/R_{\min})^2$, or $N_b \propto n^{-7/6}$. This argument suggests that the gate developed in this work might be compatible with larger GHZ-state sizes when operating at smaller internuclear separations and lower n. While we find this rough scaling argument to be instructive, we emphasize that the assumption $R_{\min} \propto R_{\times}$ requires careful scrutiny. Practically one has to consider the tunability in the atomic separation (which is often fixed in many standard optical lattice designs), as naively (i.e. without considering spaghetti) larger n is beneficial for a fixed $R_{\min}$ set by the lattice spacing. Additionally, Ref. [2] mentions using a careful choice of lattice spacing (or potentially n for a given spacing) to maximally avoid molecular resonances even when inside the "spaghetti distance". For clarity, we note that we did not engineer our system for any such suppression in this work.

Given the results presented in this work, we expect that it should be reasonably straightforward to achieve 16-atom GHZ states (the next power of 2) with these multi-qubit gates. After further technical improvements to the experiment, an investigation of optimal n (or potentially lattice spacing) could help more quantitatively clarify the largest practical GHZ-state sizes that can be achieved. Nevertheless, we agree with the referee that scaling this method to much larger GHZ state sizes is challenging. To address this in the main text, we have added a clarifying statement that we expect these multi-qubit gates to be most effective for an intermediate range of ensemble sizes (lines 126–128).

> d. Regardless, it does not seem like the GHZ infidelity is primarily arising from finite blockade, as gauged by Figure 4c if I correctly understand the configuration (although it depends on the specific atom distances). In particular, is it not interesting that here the GHZ state fidelity is only slightly decreasing with increasing N? E.g., in this figure with the same total pulse length, the 4 qubit GHZ fidelity is much better than (2 qubit GHZ fidelity)^2. What does this imply about the underlying error mechanisms that are limiting the fidelity?

The fact that the 4-atom GHZ-state fidelity is much better than the the square of the 2-atom GHZ-state fidelity mainly suggests to us that single-particle errors cannot be the dominant issue. By single-particle errors here we mean clock rotation and state detection errors. This agrees with expectation based on our characterization of such errors in various sections of the Methods.

For an ideal gate limited purely by Rydberg decay, the fidelity would be essentially unchanging with N for a fixed $N_{\max}$ gate (up to $N_{\max}$ of course). This is because in the fully blockaded limit, the effective time spent in the Rydberg state is similar for the different excitation sectors n. This ideal performance is indicated by the gray shaded area in Fig. 4c. We would naively expect many sources of technical errors, as well as imperfect blockade, to contribute to the observed trend of decreasing fidelity with N for a given $N_{\max}$; however, this is not necessarily always the case, as shown by the effect of global Rydberg detuning fluctuations in the new Extended Data Fig. 4a. Developing a better understanding of how certain types of errors affect different n is an interesting direction for future work, as it could allow for the use of varying ensemble sizes as a general diagnostic tool for multi-qubit Rydberg gates.

> e. Relatedly, as there are clearly other decoherence sources, there should be some estimation of these and how they affect this protocol. It is OK if these are rough estimates and not detailed atomic physics calculations, but e.g. is Doppler a major contribution here? What about the positional spread from the drop-recapture (with a combo of concerns on \spaghetti" and finite interaction strengths)? Practitioners and theorists will consider the authors' GHZ protocol for various applications with atom arrays, and such discussion / estimates would increase the impact of their manuscript.

We have included error modeling for the Bell-state and 4-atom GHZ-state fidelities in the new Extended Data Fig. 4b. These are mostly obtained from master equation simulations with additional Monte Carlo sampling for shot-to-shot or inhomogeneous fluctuations of Rabi frequency and detuning parameters. In both cases, our error model accounts for roughly 1/3 of the observed measurement-corrected infidelities (see Extended Data Table 2). Understanding the unaccounted for infidelity remains a critical task for evaluating the ultimate utility of these multi-qubit gates; in the new Methods section “Sources of error in GHZ-state preparation”, we have added discussion of various other improvements that can be made to address sources of error which are not included in the model.

> f. The authors state "The calculated gate duration TG scales sublinearly in Nmax, potentially enabling improved GHZ-state fidelities compared to successive application of two-qubit gates when gate errors are dominant." This is not clear, and depends on the error mechanisms (although I appreciate the use of the word "potentially"). An N-qubit GHZ state creation circuit requires N (parallelizable) gates, but $\log(N)$ depth, compared to the $N^{0.59}$ scaling found here. So , e.g., if fidelity decays with total pulse time (as is seemingly being observed here) and not e.g. total time spent in the Rydberg state, then wouldn’t the log depth GHZ circuit be better? Some discussion along these lines would be useful.

We updated the sentence to be more clear that this statement is specifically meant with Rydberg decay in mind. We did not mean to imply anywhere in the manuscript that there was evidence of infidelity scaling with “total pulse time” as opposed to “total time spent in the Rydberg state”. For the pulses used in this work, our simulations suggest that these two quantities are proportional, making it difficult to distinguish such an effect from the current data (though we do not rule out such a possibility). We suspect that the final sentence in the cascaded GHZ state section discussing Fig. 4c may have caused this confusion; given the added error modeling in the Methods, we have opted to remove this sentence.

More broadly, we agree with the referee’s comment that in the presence of other errors, it is generally unclear whether a single multi-qubit gate will out-perform a more traditional circuit such as the log depth option mentioned. For the multi-qubit gates, we expect key factors will be finite blockade, quality of radio-frequency to optical phase transduction (for instance, low-pass filtering of higher-frequency components in the modulation present for larger N as seen in Fig. 1b), and heightened sensitivity to certain errors (such as greatly increased sensitivity to Rydberg Rabi frequency as shown in the new Extended Data Fig. 4a). For a traditional 2-qubit gate circuit for preparing GHZ states, challenges associated with reconfiguring the qubits must be taken into account [5, 6, 7]; depending on the time needed for reconfiguration, idling errors such as clock scattering/decay can become significant as the system will be in an increasingly large GHZ state after each CNOT gate layer. Because these multi-qubit gates will primarily be of use for some intermediate range of sizes, to evaluate the relative performance of the strategies requires a fairly complete and accurate characterization of the errors which we currently do not have. It is also worth noting that coherent transport of neutral-atom optical clock qubits in 2D is still to be demonstrated and characterized. Both directions are certainly interesting to explore further in the future.

> g. To summarize all the above, the authors manuscript would benefit from more detailed description of the limitations of this approach to creating GHZ states. Although the authors’ protocol benefits from simplicity of global pulses, there are (a) fundamental limits on size (which the authors are approaching) given by the tradeoff between "spaghetti distance" and "Blockade range", (b) there is (as stated) a significant contrast reduction for smaller ensembles, and (c) for error mechanisms which scale with pulse length (as appears to be happening here) then naively a $\log(N)$ circuit would be better than the $N^{0.59}$ pulses found here. In contrast, the authors produce a very high fidelity Bell state, likely corresponding to a gate fidelity > 0.99, which in theory should allow them to realize GHZ states of 50 atoms. The bottom-line: when analyzing these various considerations - is it better or worse to make GHZ states in this way as compared to a conventional log depth circuit? The authors’ results and new pulse techniques still stand regardless, but clearly reporting on these aspects would be very beneficial to the readership.

We thank the referee again for their detailed feedback on these points and feel that our efforts to address these concerns having significantly strengthened the manuscript.

Regarding the "bottom-line" question, we take the following perspective. In the short term, we believe the multi-qubit gate technique has great utility for GHZ-state production, evidenced by this work achieving the largest GHZ states for long-lived, neutral-atom qubits so far. As mentioned in the Methods, we expect that 16-atom GHZ states should be straightforwardly achievable, and with additional technical work to further improve fidelities, Fig. 4f suggests that a significant gain can already be achieved. From the metrology side, there can potentially be a benefit to the relative simplicity of our state preparation procedure (for instance to reduce the number of systematics one might introduce).

Long-term, we agree that a single multi-qubit gate will not scale to much larger systems. For both GHZ-state preparation and most other tasks, a more conventional circuit approach will clearly be more scalable. Nevertheless, the circuit approach may still be sped up in certain instances by having native access to multi-qubit operations. The phase modulation technique used in this work can be generalized to perform any symmetric, diagonal gate in principle. We did not explore in detail alternative strategies for using other multi-qubit gates in GHZ-state production, but multiply controlled Z gates [8] (and through that multi-qubit Toffoli gates using local single-qubit rotations) are certainly of great general interest. Because of the minimal change needed to go from one type of gate to another, we expect that essentially all advances made in working on this GHZ-state preparing gate will translate to these other applications. Thus, we feel there is great motivation for continuing to explore and improve these multi-qubit gates, even beyond the immediate application presented in this manuscript. We have added a statement making note of these possibilities for the multi-qubit gate technique to the conclusion (line 393–396).

> `2. Related to the above, the authors should provide a brief comparison, at least qualitative, to spin squeezing for metrology (including their prior work). It would be interesting to have even a qualitative perspective on, given the results in the present manuscript, does it appear squeezing or cascaded GHZ states will be more effective for sensing in the near (and long) term?`

Whether squeezed states or (cascaded) GHZ states will become more effective will depend on many considerations. Among other things, this includes how the respective state preparation procedures continue to improve in the future, the specific metrological application of interest and robustness to noise during the interrogation. We have added a sentence to the conclusion mentioning that an interesting future direction for programmable clock systems is examining the practical utility of different entanglement strategies (line 389–393). We provide some additional discussion of some of these aspects below for squeezed and GHZ states.

Restricting to the task of achieving the best possible clock stability, squeezing might be considered more immediately useful as an improvement has been demonstrated at significantly lower instabilities [9, 10]; in this work, we have only demonstrated an improvement at a higher instability and restricted dark time for single GHZ-state sizes, and have not yet achieved improvement with the cascade. However, the next generation of few thousand qubit atom array processors [11, 12, 13] and a reasonable path toward higher fidelities suggest a promising near-term future for cascaded GHZ states; these systems could reach near state-of-the-art QPN-limited stability with maximum GHZ-state sizes between 50–100 atoms. There is also significant room for improvement in preparing spin-squeezed states, though we leave an investigation of this to future work. Outside of state preparation, factors such as clock state lifetime (due to trap scattering or spontaneous decay) and dead time effects (see Ref. [14] for analysis of this consideration for squeezed states) should also be taken into account to accurately assess the actual improvement in stability attainable with these entangled states.

More broadly, there may be various sensing applications where GHZ states can still be preferable to squeezed states (potentially even without cascade). Some of these were mentioned in the last paragraph of the GHZ state atom-laser comparison section, and in short arise for sensing applications focused on a high-bandwidth where being able to perform more measurements quickly is helpful. Another exciting longer-term opportunity is that of clock networks with remotely shared entanglement, for which protocols using GHZ-state cascades are known to optimally utilize resources [15]. Ultimately, we expect that both spin-squeezed and GHZ states will have utility in various contexts.

> `3. I ask the authors include their new pulses as some .csv or .txt supplementary files with the published version of the manuscript, if this is not already the case.`

In the published version, we will include csv files of the pulses for each point plotted in Fig. 1b.

> `4. Throughout the manuscript the authors say they are doing X(pi/2) gates. I believe the authors mean that they are doing Y(pi/2) gates, as a rotation around the Y axis is what creates a superposition |0> + |1>, whereas a rotation around the X axis creates |0> + i |1>. Although just notation, please check this and ensure consistency with the literature.`

We have double-checked the notation and believe it is all consistent. However, we realize that one confusing thing might have been the exclusion of the $Z(-\pi/2)$ rotation inherent to $\mathcal{U}$ [see Eq. (9) in the Methods] for the circuits shown in the figures. While we did this to minimize clutter, we realize that it may be misleading as applying only the CZ gate portion of $\mathcal{U}$ would require a $Y(\pi/2)$ rotation in place of one of the $X(\pi/2)$ rotations to produce the GHZ state. We have thus added in explicitly the $Z(-\pi/2)$ rotation into the circuit depiction of $\mathcal{U}$ in all figures. As shown explicitly in the methods [see Eqs. (10)–(12)], the sequence $X(\pi/2)\mathcal{U}X(\pi/2)$ does produce a GHZ state.

> `5. Technical question: the authors describe their intensity modulation when scanning phase of the AOM. Why do the authors not just double-pass their AOM? In order to preserve optical power? Furthermore, are the 5-10% intensity modulations not compensated at all? Would this not give a massive miscalibration error? How are the pulses being calibrated, and can miscalibration simply be the source of the fidelity degradation with GHZ size? Please comment on this.`

We agree that double-passing the AOM, or using two AOMs in a tandem configuration, should greatly suppress the issue with transduction of phase modulation to amplitude modulation. Maintaining a sufficiently large Rabi frequency was indeed the key consideration. We hope to upgrade this aspect of the experiment in the future and have noted this potential improvement in the new Methods section (line 1236–1238).

The amplitude modulation is definitely a source of error for the applied gates. Beyond trying to perform regular alignments of the fiber, we did not apply any compensation of this effect. Feedforward was complicated since the modulation would visibly change over time as the beam pointing was drifting; feedback was not possible due to the loop bandwidth compared to the gate speeds.

In terms of calibration, we began by aligning the fiber as best as possible to minimize the modulation. After this alignment, we would then calibrate the Rabi frequency by performing Rabi oscillations. For the two-qubit gate, we checked the relative pulse amplitude with and without modulation on a photodiode and observed a difference between 0.25–0.5%; note that fractional change in Rabi frequency is approximately half of that. We have not explicitly checked how much this changes for larger $N_{\max}$. We also note that the Rabi frequency calibration procedure is likely less accurate than 0.5%, which could certainly be a limitation. We have added some comments regarding this in the new Methods section (line 1225–1231).

> `6. Can the authors describe how much power is on the atoms and what the size of the Rydberg beams are, for their 4 MHz rabi frequencies?`

There is roughly 150 mW of UV light incident on the atoms, and the beam is focused to an elliptical spot of roughly 240 μm × 5 μm ($1/e^2$ radii) [16].

> `7. With respect to metrology, this work provides very valuable proof-of-principle demonstrations. However, I think the presentation can be improved and it can be clarified in multiple places that these results are proof-of-principle.`

> a. The authors present results beyond SQL. However, the GHZ state coherence times are very short, and are not even limited by laser phase noise but by unrelated global correlated magnetic field noise. As such the authors measure at a short dark time of 3 ms. Although the GHZ outperforms the CSS for this particular dark time, can the authors justify that this is truly ‘‘below-SQL performance"? Since, of course, one should measure with a dark time ∼ coherence time, and the GHZ state has quite reduced coherence times. My concern about this point can be fixed by appropriate wording modifications in the manuscript.

We have updated wording in the abstract (3rd to last sentence) and intro (lines 68, 75–78) to further emphasize that the “below-SQL performance” applies only for a restricted set of sufficiently short dark times, and not the optimal achievable stability.

> b. The inset for Fig 3b is confusing. It initially appears that the authors are saying the CSS and SQL are scaling with the same slope as the Heisenberg limit. I suggest making the presentation clearer. Relatedly, as plotted it appears that the grey circles are approaching the HL scaling, and not the other way around that the HL scaling is modified to be the gray circles.

We have adjusted the legend label to make the meaning more transparent. We have also updated the gray circles to gray points with arrows pointing from the HL scaling line; we hope this clarifies that it is the HL and not the data that is being modified.

> c. The statement "These gains can be practically harnessed for a restricted class of problems...", is this still true here given that the reduction in coherence time is not even coming from laser phase noise, but from unrelated global magnetic field noise?

The intent of this sentence was not to imply anything specific about the current experimental situation, but rather to point out that GHZ states of a fixed size do have utility despite failing to improve the optimal attainable stability (as discussed in the remainder of that paragraph). To make this more clear, we have reworded “These gains...” to “The metrological gain of a single GHZ-state size...”. Indeed to practically benefit from the gain in stability requires the signal being detected to be dominant. So the GHZ states demonstrated here could help if a much less stable clock laser was used (of course still under a dark time restriction), or to measure time-varying signals at a certain bandwidth where it could be useful to make more measurements at intervals much shorter than the coherence time. It is also worth noting that the single atom coherence time in our setup can be improved to the second-scale by reducing the magnetic field [17] or by reducing noise in the current supply used to generate the field.

> d. If I understand correctly, the authors operate with up to 30 atoms at a time, but then postprocess data to include multiple runs as "simultaneous ensembles" and get to a "total atom number" of 118 atoms. Since these are proof-of-principle experiments, and this bootstrapping allows for probing physics of the protocol, I think it is ok. However, "total atom number" could mislead readers and should be re-worded to make it clear that this is e.g. "postprocessed atom number". Technically this also means that previous experiments where people made variable-sized GHZ states but in different shots could also be postprocessed to do this cascaded GHZ state protocol. I think it’s fine, but the authors should word more carefully.

We have changed the wording to “bootstrapped total atom number” in the relevant locations to make this point more clear. The referee’s assessment about showing this proof-of-principle with separately prepared GHZ states of different sizes is true, though we would note one point. By still postprocessing on data from a cascade, the results in Fig. 4f more accurately benchmark the achievable cascade performance given our currently available state preparation protocols and fidelities. We could have alternatively showed results combining data from Figs. 1c and 2b (along with an additional CSS Ramsey fringe) where each GHZ-state size is prepared with a different gate. This would yield significantly better performance, but we would be unable to produce such a cascade without a reasonably significant upgrade to the experimental architecture. This is why we chose to perform the postprocessing on an actually simultaneously prepared cascade.

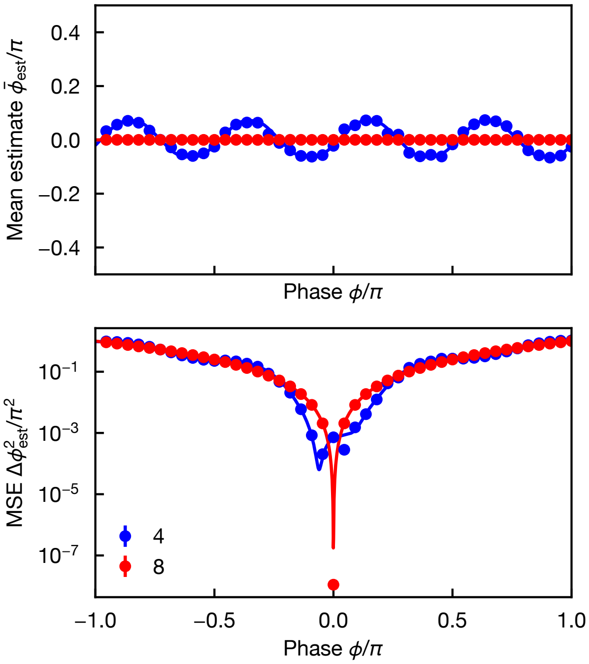

Figure 1: Mean estimator and mean-squared error for individual GHZ-state sizes. Blue is for $N = 4$ and red is for $N = 8$-atom GHZ states. The markers are the bootstrapped experimental data and the lines are based on the fitted model.

> e. The theoretical dynamic range and mean-square error of Nk = 8 are plotted in Fig 4d and 4e. However, don't the authors have the data to just experimentally plot these? Is there a reason that they did not? This is not necessary but would be interesting to see experimental data here.

The choice to show the theoretical range and not experimental data was purely to reduce visual clutter due to the high density of traces and labels in those plots. We have added these traces for $N = 4, 8$ GHZ states below for the referee. The shapes of these curves will vary significantly depending on the assumed prior width (or particularly the ratio of the width to the dynamic range); here it was taken to be $\pi/6$ as in the manuscript.

> f. Is 4f assuming something particular about the laser phase noise distribution? If so, please state it.

Fig. 4f assumes the atom-laser phase prior is a Gaussian with standard deviation of $\pi/6$ (line 345–350). This assumes that the atom-laser phase diffuses outward in a normal manner (line 317–320), but this is essentially always satisfied.

Response to referee 2

> Summary In their manuscript "Multi-qubit gates and 'Schrödinger cat' states in an optical clock" the authors demonstrate their advance towards realistic metrological gain of using entanglement in atomic optical clocks with tweezer arrays. The authors provide several suites of characterization of their efforts, including the generation of GHZ-states up to size 9, a cascade of GHZ-states of increasing size, and detailed characterization of the phase estimation properties of both of these. The authors supplement their main text with a comprehensive methods section that enables a deeper understanding of the material.
> High-level feedback My overall assessment of the manuscript in its current state is that it constitutes an impressive effort, written in an accessible and at the same time accurate fashion. The figures are both informative and beautiful, though at times I feel like in print a magnifying glass will be required. The authors have successfully navigated many of the pitfalls about statements in quantum metrology, which I greatly appreciate. The manuscript has clearly been diligently proof-read; I can find only a single potential typo. This work has been made possible by the application of technical and theoretical advances of recent years and is now pushing the field of quantum metrology further. I do have a number of comments that I believe will improve the clarity of the manuscript, but if it were published as is today it would still constitute one of the better Nature papers I have ever read. It was a pleasure reading it and I congratulate everyone involved.

We thank the referee for their positive characterization of the manuscript and their detailed feedback.

> Detailed feedback
> Abstract: There are two things that come to mind in the abstract. Firstly "sufficiently short dark times" is perhaps the main issue with quantum-enhanced phase estimation. You win linearly in Ramsey time while you win *at best* linearly with number of probes, so in general you lose when using entanglement if you account for increase in duty cycle etc. . While the authors address this in the manuscript to an extent, it is an important point that ought to be highlighted - perhaps in the discussion section rather than the abstract. The second thing is the jump from "perform below SQL" for up to 4 qubits and then moving on to "A key challenge". This is also something that I struggled with in the manuscript main text and I will get to that at the appropriate section. However, I think it is not sufficiently clear how the two sentences immediately following each other are related. One might wonder why below SQL is not enough, for example. See comments to Fig. 4.

We have changed the "key challenge" sentence in the abstract to be more explicit that what a single GHZ-state size fails to improve is the clock stability after optimizing for dark time. We hope this makes clear the connection with the preceding sentence, where SQL can only be beat for a fixed and sufficiently short dark time.

We would push back on the referee's general assessment that "you lose when using entanglement". Regarding the specific point about duty cycle, we note that the time for preparing these GHZ states can be much shorter than the Ramsey time in principle, and the atom array architecture is well-suited for parallel operation where new GHZ states are prepared during an ongoing interrogation which would entirely eliminate this issue. We also note that the linear improvement with Ramsey time is not always the case, and for ideal zero dead time operation the improvement would be square root in the Ramsey time. Independent of these points, the primary purpose of Fig. 4 is to demonstrate that the reduction of Ramsey time for GHZ states (which is due to their smaller dynamic range when limited by atom-laser coherence) can be overcome by using a cascade of GHZ states. This will allow one to win from the entanglement for larger cascades without having to reduce the Ramsey time at all (up to a point where another decoherence source becomes limiting, such as clock-state lifetime).

> Introduction: I had expected to see a reference to the Oxford paper on entangled clocks in the introduction. This is something the authors may consider adding.

This was certainly an oversight, and we have now added this reference to the conclusion where we discuss remotely-entangled clock networks (line 376).

> l. 15: Here the Heisenberg limit is equated with the 1/N scaling and it is explicitly stated that this is the fundamental bound given by quantum theory. If that is so, then we cannot explain "super-Heisenberg" scaling within quantum theory, which is of course not the case. Instead, Heisenberg scaling for linear observables is limited to 1/N and what people call "super-Heisenberg" is just Heisenberg for a nonlinear observable. I would advise adding the single word "linear" to stay correct outside of the most-common case of linear observable (here σ_z) and stand against the \super-Heisenberg" misnomer.

We thank the referee for pointing this out and have added a clarification to the text that $1/N$ applies for linear observables (line 20–21).

> l. 24: One could expand the list with use-cases of real-life application from the somewhat arcane gravity and dark matter to life science with "Quantum-enhanced nonlinear microscopy" also in Nature, in particular in view of the wider readership of Nature.

We have added the suggested reference to this sentence (line 25).

> l. 43: I am unsure what exactly ‘scalable’ means here. In particular I am not sure if it is true if I just use my own interpretation without context due to the distance scaling of the Rydberg blockade. Or is this meant more generally that in principle entanglement can assist at this front? Overall, I am not sure where this sentence is going.

We see that this may have been misleading and have removed the word ‘scalable’. We did not mean to imply that the multi-qubit technique presented here is scalable, largely for the issue of the finite Rydberg blockade range as the referee mentions.

The intent of this sentence is to point out that on the path to merging the full capabilities of atom-array quantum processors and optical lattice clocks, the next goal for tweezer clocks is demonstration of high-fidelity entangling gates. These gates can then allow for systematic production of (large) entangled states that can improve clock metrology, which is the underlying motivation for the manuscript.

> l. 47: I am not sure what "macroscopically distinct" means in this context. I feel like you mean "mutually orthogonal"? As in ‘x’ or ‘not x’. Nothing here is macroscopic?

We agree with the referee that the systems sizes are far from traditionally macroscopic (though this is true of all GHZ states that have ever been experimentally produced). We note, however, that the intent of this sentence was to introduce GHZ states in a broader context. The term “macroscopically distinct” is commonly used throughout the literature to describe cat states and GHZ states; for instance, see Refs. [18, 19]. We agree that “mutually orthogonal”, in the sense that each individual spin is in orthogonal states between the two components of the superposition, captures a similar essence. However our impression is that “macroscopically distinct” is the more familiar term for most readers due to its connection with the well-known “Schrödinger’s cat” idea. Additionally, there are formal measures characterizing this notion of macroscopic distinctness (or more generally macroscopic superposition) [20] for which GHZ states exactly satisfy the conditions of “macroscopic”-ness exactly for all N (as opposed to asymptotically). We have added a reference to this, but left the terminology as is for these reasons.

> `l. 50: I was surprised by this "thus". Is the fact that the GHZ-state has N-times faster oscillation in parity actually the reason for saturating the HL? I thought both facts are true, but not that one causes the other.`

We have changed "thus" to "and" in order to avoid any confusion. Here we make a few remarks to clarify the connection.

For exactly saturating the minimum possible variance in parameter estimation, these two properties are equivalent. The quantum Fisher information (QFI), which for pure states is equal to the variance of the encoding operator/Hamiltonian (collective spin along z), sets the minimal parameter estimation variance via the quantum Cramér-Rao bound [21]. The state with maximal variance in the encoding Hamiltonian (or maximal QFI) is of course also the state which accumulates phase maximally (superposition of eigenstates with extremal eigenvalues). The specific "N-times faster" reflects the difference of N between extremal values of an N-qubit collective spin operator. The equivalence of these points means that actually the "thus" goes both ways (in the sense of using "thus" to mean "A implies B").

For saturating the Heisenberg scaling behavior instead of the exact bound, there are other types of states which can achieve this such as squeezed states from two-axis twisting [22]; these states reach the Heisenberg scaling limit due to noise reduction as opposed to faster phase accumulation. In general the Heisenberg scaling can be caused by different properties for different classes of entangled states; it is in this sense that we were originally using the word "thus".

> `l. 67: Here the authors mention that this is the first time that GHZ states are used for below-SQL performance. I am wondering whether they are being a very 'technically correct' or missing references. Both [3] and [63] could argue that they have done so. [3] has at admittedly extremely short Ramsey times and self-referenced which one may reasonably have reservations against. [63] has shown at least parts of the performance below the SQL. Of course, neither have done GHZ but instead squeezed states or optimal probe states, but the GHZ is just a squeezed state for a particular squeezing parameter.`

We agree with the referee that below-SQL performance has been achieved before. In addition to the references mentioned, we demonstrated a clock instability (for a differential atom-atom comparison) below the SQL with squeezed states on this same apparatus previously in Ref. [10]. However, the use of GHZ states, as opposed to other types of states, for below-SQL performance is the critical aspect of this claim. We believe the distinction is important and explain further below.

The referee states that a GHZ state is "just a squeezed state for a particular squeezing parameter". We suspect that the referee might be referring to the connection between squeezed states and GHZ states through the canonical one-axis twisting (OAT) Hamiltonian for generating spin-squeezing; specifically, that the "squeezing parameter" being referred to is the time at which OAT produces the GHZ state. To define a spin-squeezed state, a separate number, which is traditionally called the squeezing parameter, is used to characterize the reduction of projection noise along a certain spin measurement axis; if this number is < 1, then a state is said to be squeezed. One widely used squeezing parameter is that of Wineland [23, 24, 22, 25]

$$\xi_R^2 = \frac{N\left(\Delta J_\perp^2\right)}{\langle J_{\vec{s}}\rangle^2}, \tag{1}$$

The Wineland squeezing parameter ξ_R^2 quantifies the variance reduction $N\Delta J_\perp^2$ compared to N uncorrelated particles along a direction perpendicular to the mean spin direction $\vec{s}$, and is additionally normalized by the mean spin length $\langle J_{\vec{s}}\rangle$. It is clear that for GHZ states ($N > 2$), the vanishing mean spin length $\langle J_{\vec{s}}\rangle = 0$ causes the Wineland squeezing parameter to diverge [22], and thus should not be considered a type of squeezed state.

More generally, the distinction between squeezing and other forms of metrologically useful entanglement can be made by considering the difference between ξ_R^2 and the metrological gain expected from the QFI. The difference is well illustrated by OAT where ξ_R^2 is only < 1 during the early time dynamics, while the metrological gain continues to improve until the GHZ state is produced [25]. The reason is that squeezing focuses on collective spin measurements while the QFI is optimized over all possible measurements. This is not only a formal distinction, but also has major practical consequences for experiments. Many systems only have access to collective spin observables, and thus only have access to the metrological gain from squeezing. Our work makes use of microscopic single-atom detection which enables the parity measurements necessary for GHZ states; other decoding protocols have been explored for other types of entangled states, such as over-squeezed and variationally optimized states as in the references mentioned by the referee. We have left the claim as is (up to a qualifying clause about Ramsey time restrictions related to other comments) for these reasons.

> `Fig. 1: An indication that it is an optical qubit would be good, in particular since some other publications insist on stressing this. In the caption either I cannot parse the \form from" or it is a typo. You mean the \shape from" or is it \from" twice? The feature sizes from the fluorescence image are going to challenge any printer and researcher above the age of 25 in print. As for the circuit: Either you use some definition I have never seen or this circuit has one typo. This thing makes |01> + |10> when I do it by hand, matrix multiplication or in qiskit. If you start in |11> it would be the GHZ or if you have either of the X-rotations with a negative sign. Matters not for what you want to show, of course, but you write explicitly it is |00> + |11>.`

We have changed “form from” to “implementation from” to avoid this confusion; indeed we meant the functional form of the phase modulation.

Regarding the circuit, the difference has to do with the CZ gate implementation given in Ref. [8] which we used. Compared to the standard definition of the CZ gate, their implementation effectively has an additional $Z(\pi)$ rotation; it applies the -1 to the $|00\rangle$ state instead of the $|11\rangle$ state. Originally we followed the notation for the circuit defined in Ref. [8]; however, because we explicitly use the standard definition of the CZ gate elsewhere in the manuscript, we have changed the notation to $X(-\pi/4)$ as suggested by the referee.

> `l. 76: The authors use a Ramsey time of 3 ms, which is substantially below their coherence time. As discussed above, Ramsey time wins linearly in fractional instability so should always be as long as possible. I do not know what the rule of thumb is for how much below your coherence time you should stay before the Ramsey time starts to hurt more than aid. Is the dark time chosen such that at N = 4 you are just below this rule? If so, state that? If not, why are you at 1% of your coherence time?`

Indeed a longer dark time could have been used to obtain a lower instability, and general guidelines for choosing dark time when limited by atom-laser coherence can be found in Ref. [26]. However as mentioned in the manuscript and by the referee in an earlier comment, GHZ states of a single size are not particularly useful for improving the absolute best stability of a clock. Their main utility is in situations where there is a separate dark time or bandwidth restriction apart from atom-laser coherence. For this reason, we simply chose a conservative value for the dark time which is well within the 4-atom GHZ-state coherence time; we have added a clarifying statement in the text (line 230–232).

> GHZ design and GHZ-state preparation: l. 103: This entire sentence was very confusing, because n is used inconsistently. First $\backslash\sqrt{n}$-enhanced", so n is clearly a number. Then n is the sum of projectors, so an operator. Can one of them get a hat? Overall, I was repeatedly confused because I forgot what n and N are. Perhaps the operators can get a different symbol or some such.

We agree that the notation was confusing and have added the hat notation for operators across the manuscript. Locations without a hat (such as the one mentioned) now indicate eigenvalues of the corresponding operator.

> l. 107: I was perplexed for a while as to what the Rabi trajectories are. When I hear trajectories in the context of gates I think of phase space trajectories, which is probably not what is meant here. Instead, do the authors mean that the different sub-ensembles see different laser parameters because they are spatially separated and therefore they use a pulse that is insensitive to fluctuations of these parameters? That is two ensembles at different locations see, say, different Rabi rate so you chose a pulse that is robust against Rabi rate changes to implement your 2π pulse up and down to |r>?

We have adjusted the wording to "blockaded Rabi oscillations for different n" to be more explicit. To explain in more detail here (also see Methods section "Optimal control for multi-qubit gates"), every computational state undergoes a separate blockaded Rabi oscillation in an effective two-level system with a W-state of a single Rydberg atom among the n-atoms initially in $|1\rangle$. Because of the Rydberg blockade, the Rabi frequency of this oscillation is increased by a factor of $\sqrt{n}$. The trajectories we are referring to are those of the effective two-level Bloch vector dynamics on the Bloch sphere for different n. A single pulse will produce $N+1$ different trajectories on the Bloch sphere (for $n = 0, 1, \ldots N$, though $n = 0$ is trivial as there is no Rydberg excitation). The trajectories start and end at the south pole of the Bloch sphere (the computational state), and accumulate a phase corresponding to the desired gate. An example of these trajectories for the CZ gate is shown in Fig. 3b of Ref. [27].

> l. 136: Again, I think that makes the wrong Bell state unless it is $-\pi/4$ or you start in |1> or it is $-\pi/2$ in the start, no?

We have changed the $\pi/4 \to -\pi/4$ at all relevant locations (see previous comment regarding circuit in Fig. 1).

> l. 172 / Fig. 2: I was perplexed by this graph until I realized grey points are not grey markers. Perhaps call out the dashed line or give one of those a different colour.

We have added the call out to the dotted line.

> GHZ-state atom-laser comparison: l. 226: Here again the question arises why it is 3 ms and not more, as discussed above.

We have added a clarifying statement (see previous comment).

> Fig. 3: The fluorescence images are very small. It took me a fair bit to parse the inset. I concluded that this is \how much better am I than", but perhaps this could be said more directly. It is all there, I know, but I still had trouble with it.

We have increased the size of the image by 25% and changed the legend label to make the meaning more transparent.

> Cascaded GHZ-state phase estimation: l. 347: This section makes no mention of what T is used. Is it still 3 ms? If so, that should be stated somewhere in the section or figures.

For the experimental data, the measurements in Fig. 4d-f are all performed at zero dark time $T = 0$, and are a bootstrapped reanalysis of the exact same data as shown in Fig. 4b; we have reworded a sentence to make this point more clear (line 309–314). The idea is that this characterization of the phase estimation properties at zero dark time (Figs. 4d-e) can be used to project the performance in actual clock operation at a non-zero dark time (Fig. 4f) by taking the atom-laser phase to be a random variable with a Gaussian distribution (line 343–349 and Methods section "Effective measurement uncertainty for frequency estimation); the width of the distribution would be determined by a specific dark time and laser noise model.

> Fig. 4: Looking at sub-figure f the question arises how both abstract and the preceding section claim that the authors have beat the SQL but then here and in the main text of this section they say that the scheme is still not at the SQL by a 2 dB margin. Is the statement that with the short Ramsey times above you beat the SQL, but in a realistic scenario you would not with a single GHZ be able to due to the limited dynamic range? If so, then this section would be about using a larger Ramsey time, but that is not stated. If not and you are still at 3 ms then this section would be an \in principle we would need to" and then the limit would be something else. For example, because it is harder to make several ensembles rather than one. Is this the case? Either way, that bit was unclear to me from the very start and the presentation should be sharpened to chisel out this point more clearly.

There are a couple main reasons that the cascade currently falls short of the SQL, in contrast to the individual GHZ-state sizes.

The first is that the cascade has more stringent requirements for beating the SQL. The cascade requires an overhead of resources to overcome the dynamic range limitation of an individual GHZ-state size, and thus one should not expect that the fidelity requirements for improvement are the same. The overhead can readily be seen in the gray points of Fig. 4f, which show the result for a cascade with perfect contrasts; the empirical $\pi^2 \ln(N_{\text{tot}})/N_{\text{tot}}$ line matching this ideal cascade lies significantly above $1/N_K$, which is the naive Heisenberg limit for maximum GHZ-state size. Some analysis of the increased threshold for improvement is discussed in the Methods of Ref. [7].

Additionally, it is also indeed "harder" for us to make several ensembles of different sizes in the following sense. Since we use only a single global pulse designed for the largest GHZ-state size (see Fig. 4a and line 291), the fidelity of the smaller ensembles tends to be reduced from the best we can produce when preparing only a single size. This is shown in Fig. 4c, and this limitation of our scheme is mentioned on line 306–308; we have added an additional sentence mentioning this issue in the discussion of Fig. 4f (line 354–356). The importance of this effect for the Fig. 4f data is apparent from the $N_K = 1$ CSS point; without applying the Rydberg pulse, our typical Ramsey contrast of ≥ 0.995 would be within 0.1dB of SQL. As mentioned in the conclusion (line 380–384), there are various possibilities for overcoming this challenge.

> Conclusion: l. 387: I just wanted to point out that VQA on hardware is typically quite demanding in terms of number of shots, which for cold atoms are the costliest of all platforms. If the many iterations typically required on the platform are not run substantially faster than drift rates or if the fluctuations during the optimization are sufficiently large then this is going to be a fruitless endeavour. Which is not to say that it cannot be done or the authors should not mention it, only that it is very hard to do well.
> I would have liked a short discussion on the topic of Ramsey times in the seconds for optical clocks with uncorrelated particles versus entanglement-enhanced ones given that the coherence time falls quickly with ensemble size.

Our results in Fig. 4 demonstrate how GHZ-state cascades can recover the dynamic range of unentangled particles. In current state-of-the-art optical clocks, laser noise is the dominant limitation to the coherence time, and thus the GHZ-state coherence time reduction is expected to result from their reduced dynamic range. Because the cascade fully recovers the dynamic range, ideally there will be no dependence of the Ramsey time on the maximum GHZ-state size. This does not scale indefinitely as eventually the clock-state lifetime will instead become the limiting factor on the coherence time.

> Methods: l. 1015: The sequence here is correct but it is not the one in the figure.

We removed certain global Z rotations in the figures for the sake of clarity (previously stated in the caption of Fig. 2). One of these rotations is the $Z(\alpha_c)$ rotation referenced here; it's purpose is to compensate for a $Z(-\alpha_c)$ rotation caused by details of the Rydberg excitation (mentioned in line 1036–1039). This is commonly the case for Rydberg gates and is standard to exclude in circuit representations; we mention it here simply to be explicit about the experimental details.

We also excluded the $Z(-\pi/2)$ rotation inherent to the form of $\mathcal{U}$. However, we see that this could be confusing as the circuit would require a $Y(\pi/2)$ pulse in place of one of the $X(\pi/2)$ pulses to produce the GHZ state if only the CZ gates are applied. For this reason, we have now added the $Z(-\pi/2)$ into the circuit depiction of the multi-qubit gate $\mathcal{U}$.

> l. 1022: That is the one in the figure and here the authors correctly say that it produces the (other) Bell state. Generally, I find calling one of them "the Bell state" confusing, since there are four and the GHZ is one of them.

We believe this confusion is remedied by the previous comments regarding the discrepancy in definition of the CZ gate and $X(\pi/4)$ rotation. To further clarify, by Bell state we always mean the $N = 2$ version of Eq. (1) $(|00\rangle + |11\rangle)/\sqrt{2}$ and never $(|01\rangle \pm |10\rangle)/\sqrt{2}$. It is also worth noting that we always use the Bell state terminology to refer to $N = 2$ and GHZ state for $N \geq 3$.

> Eq. 8 / l. 1055: You write |$+_x$> in the equation but |$+_y$> in the text. As far as I can see, the text is correct, otherwise you are missing a final z-rotation.

As shown in Eq. (9), our multi-qubit gate $\mathcal{U}$ performs an additional $\pi/2$-rotation which maps $|+_y\rangle \to |+_x\rangle$. It is standard to define the graph state as CZ gates applied to the $|+_x\rangle$ state; our gate $\mathcal{U}$ rotates $|+_y\rangle \to |+_x\rangle$ in addition to applying the CZ gates. Explictly, for the fully connected graph $G = (V, E)$ we have

$$|G\rangle = \prod_{(a,b)\in E} \mathcal{U}_{\mathrm{CZ}}^{(a,b)} |+_x\rangle^{\otimes V} = e^{-iN\pi/4}\mathcal{U}\,|+_y\rangle^{\otimes V} . \tag{2}$$

As per a previous comment, we have now explicitly added the $Z(-\pi/2)$ rotation to the circuit depiction of $\mathcal{U}$ in the figures.

> l. 1063: You do not mean all of U is just a number, correct? You mean the prefactor before the operator exponential, no?

Similar to the case of n, here we were referring to the eigenvalues of $\mathcal{U}$. Per that previous comment, we have now included hat notation for operators to make clear that corresponding symbols without a hat indicate eigenvalues.

> l. 1107: The authors say that the rearrangement success rate varies between 85 - 98%. Do I know before measurement? What does this mean for the duty cycle? I believe the authors allude several times to there being fewer than expected atoms in some GHZ states, which would mean you do not know beforehand and just live with it being worse than anticipated?

Up to imaging infidelity and loss as described in the "State detection" Methods section, we know the exact arrangement of atoms before performing and clock or Rydberg pulses. For most of the results, we post-select the data for observing a specific fill within an ensemble (i.e. the 4-atom GHZ-state fidelity is inferred from ensembles that we observe to have 4 atoms before performing the GHZ-state preparation protocol). Thus, the post-selected success rate is much higher than this range and we do not expect it to be a dominant source of the inferred GHZ-state infidelities. It does not change the duty cycle of the experiment (we do not perform additional rounds of rearrangement to fill in missing atoms), but does reduce the rate of data acquisition.

The only data where we do not post-select is Fig. 3 (but we do still know how many atoms were initially in each ensemble). Unlike other data performed at zero dark time where results across different repetitions of the experiment can be averaged (because there is no integration of the stochastically varying atom-laser detuning), clock operation with nonzero dark time requires single-shot estimates of the atom-laser detuning. Since there are only a few ensembles on each shot, there is a significant increase in projection noise from throwing out (i.e. post-selecting) ensembles with a missing atom. Because our protocol still produces the GHZ state for imperfectly filled ensembles, we instead include them in our analysis (described in detail in the

Methods section GHZ-state stability in atom-laser comparison). Nevertheless, this imperfect rearrangement and subsequent averaging over smaller ensembles does degrade the metrological gain (essentially reducing the average sensitivity), and is a significant contribution to the "expected performance" above the Heisenberg limit shown in the inset of Fig. 3 (arrows pointing to gray circles). Improving and maintaining a high rearrangement success rate will be important for clock operation with larger GHZ-state cascades.

Additional changes

We thank both referees again for their extensive and insightful feedback. We have made a few other modifications to the manuscript unrelated to those comments which we describe here.

1. In Fig. 2b, there was an error in the definition of the phase (x-axis) in the parity contrast data analysis. This consists of a negative sign and a left-right flip of the data. This has been corrected, but does not change any of the fitted parity contrasts.
2. There was a discrepancy in the definition of the parity between the main text and the Methods. The two definitions differed by a minus sign for odd ensembles sizes N only. We have now made the definitions all consistent throughout and propagated the changes through the analysis. Again this does affect the inferred parity contrasts.
3. We modified the final sentence in the "Clock and Rydberg coherence" section of the Methods to include our estimate of the Doppler dephasing contribution (line 1001–1004).
4. Due to length restrictions, we have removed or reduced wording in certain places to accommodate for the added discussion in other locations. This includes removing a sentence about possible error sources for the fidelities in Fig. 4c (mentioned in response to referee 1) and removing a sentence about the asymmetry in the parity oscillation for the experimental cascaded data in Fig. 4d.
5. To comply with reference limits, we have removed around 9 references in the main text, many of which have been moved to the Methods. Including references added in response to the referee comments, the total number of references has been reduced from 68 to 62 in the main text; the total number of references overall has increased from 79 to 82.

References

[1] Sven Jandura and Guido Pupillo. Time-Optimal Two- and Three-Qubit Gates for Rydberg Atoms. *Quantum*, 6:712, May 2022.

[2] Andrei Derevianko, Péter Kómár, Turker Topcu, Ronen M. Kroeze, and Mikhail D. Lukin. Effects of molecular resonances on rydberg blockade. *Phys. Rev. A*, 92:063419, Dec 2015.

[3] Sebastian Weber, Christoph Tresp, Henri Menke, Alban Urvoy, Ofer Firstenberg, Hans Peter Büchler, and Sebastian Hofferberth. Tutorial: Calculation of Rydberg interaction potentials. *J. Phys. B: At. Mol. Opt. Phys.*, 50(13):133001, 2017.

[4] Robert Löw, Hendrik Weimer, Johannes Nipper, Jonathan B Balewski, Björn Butscher, Hans Peter Büchler, and Tilman Pfau. An experimental and theoretical guide to strongly interacting rydberg gases. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 45(11):113001, may 2012.

[5] Dolev Bluvstein, Harry Levine, Giulia Semeghini, Tout T Wang, Sepehr Ebadi, Marcin Kalinowski, Alexander Keesling, Nishad Maskara, Hannes Pichler, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin. A quantum processor based on coherent transport of entangled atom arrays. *Nature*, 604(7906):451–456, 2022.

[6] Dolev Bluvstein, Simon J Evered, Alexandra A Geim, Sophie H Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, et al. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626:58–65, 2024.

[7] Ran Finkelstein, Richard Bing-Shiun Tsai, Xiangkai Sun, Pascal Scholl, Su Direkci, Tuvia Gefen, Joonhee Choi, Adam L. Shaw, and Manuel Endres. Universal quantum operations and ancilla-based readout for tweezer clocks, 2024.

[8] Simon J Evered, Dolev Bluvstein, Marcin Kalinowski, Sepehr Ebadi, Tom Manovitz, Hengyun Zhou, Sophie H Li, Alexandra A Geim, Tout T Wang, Nishad Maskara, et al. High-fidelity parallel entangling gates on a neutral atom quantum computer. *Nature*, 622:268–272, 2023.

[9] John M Robinson, Maya Miklos, Yee Ming Tso, Colin J Kennedy, Dhruv Kedar, James K Thompson, and Jun Ye. Direct comparison of two spin-squeezed optical clock ensembles at the 10^{-17} level. *Nat. Phys.*, 20:208–213, 2024.

[10] William J Eckner, Nelson Darkwah Oppong, Alec Cao, Aaron W Young, William R Milner, John M Robinson, Jun Ye, and Adam M Kaufman. Realizing spin squeezing with Rydberg interactions in an optical clock. *Nature*, 621(7980):734–739, 2023.

[11] M. A. Norcia, H. Kim, W. B. Cairncross, M. Stone, A. Ryou, M. Jaffe, M. O. Brown, K. Barnes, P. Battaglino, T. C. Bohdanowicz, et al. Iterative assembly of ^{171}Yb atom arrays in cavity-enhanced optical lattices, 2024.

[12] Flavien Gyger, Maximilian Ammenwerth, Renhao Tao, Hendrik Timme, Stepan Snigirev, Immanuel Bloch, and Johannes Zeiher. Continuous operation of large-scale atom arrays in optical lattices, 2024.

[13] Hannah J. Manetsch, Gyohei Nomura, Elie Bataille, Kon H. Leung, Xudong Lv, and Manuel Endres. A tweezer array with 6100 highly coherent atomic qubits, 2024.

[14] Marius Schulte, Christian Lisdat, Piet O Schmidt, Uwe Sterr, and Klemens Hammerer. Prospects and challenges for squeezing-enhanced optical atomic clocks. *Nat. Commun.*, 11(1):5955, 2020.

[15] Peter Komar, Eric M Kessler, Michael Bishof, Liang Jiang, Anders S Sørensen, Jun Ye, and Mikhail D Lukin. A quantum network of clocks. *Nat. Phys.*, 10(8):582–587, 2014.

[16] Nathan Schine, Aaron W Young, William J Eckner, Michael J Martin, and Adam M Kaufman. Long-lived Bell states in an array of optical clock qubits. *Nat. Phys.*, 18(9):1067–1073, 2022.

[17] Matthew A. Norcia, Aaron W. Young, William J. Eckner, Eric Oelker, Jun Ye, and Adam M. Kaufman. Seconds-scale coherence on an optical clock transition in a tweezer array. *Science*, 366(6461):93–97, 2019.

[18] A. Omran, H. Levine, A. Keesling, G. Semeghini, T. T. Wang, S. Ebadi, H. Bernien, A. S. Zibrov, H. Pichler, S. Choi, et al. Generation and manipulation of Schrödinger cat states in Rydberg atom arrays. *Science*, 365(6453):570–574, 2019.

[19] Zehang Bao, Shibo Xu, Zixuan Song, Ke Wang, Liang Xiang, Zitian Zhu, Jiachen Chen, Feitong Jin, Xuhao Zhu, Yu Gao, et al. Schrödinger cats growing up to 60 qubits and dancing in a cat scar enforced discrete time crystal, 2024.

[20] Florian Fröwis and Wolfgang Dür. Measures of macroscopicity for quantum spin systems. *New J. Phys.*, 14(9):093039, sep 2012.

[21] Géza Tóth and Iagoba Apellaniz. Quantum metrology from a quantum information science perspective. *J. Phys. A*, 47(42):424006, 2014.

[22] Jian Ma, Xiaoguang Wang, C.P. Sun, and Franco Nori. Quantum spin squeezing. *Phys. Rep.*, 509(2):89–165, 2011.

[23] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen. Spin squeezing and reduced quantum noise in spectroscopy. *Phys. Rev. A*, 46:R6797–R6800, Dec 1992.

[24] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. J. Heinzen. Squeezed atomic states and projection noise in spectroscopy. *Phys. Rev. A*, 50:67–88, Jul 1994.

[25] Luca Pezzè, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein. Quantum metrology with nonclassical states of atomic ensembles. *Rev. Mod. Phys.*, 90:035005, Sep 2018.

[26] Ian D Leroux, Nils Scharnhorst, Stephan Hannig, Johannes Kramer, Lennart Pelzer, Mariia Stepanova, and Piet O Schmidt. On-line estimation of local oscillator noise and optimisation of servo parameters in atomic clocks. *Metrologia*, 54(3):307, mar 2017.

[27] Shuo Ma, Genyue Liu, Pai Peng, Bichen Zhang, Sven Jandura, Jahan Claes, Alex P Burgers, Guido Pupillo, Shruti Puri, and Jeff D Thompson. High-fidelity gates with mid-circuit erasure conversion in a metastable neutral atom qubit. *Nature*, 622:279–284, 2023.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have addressed all of my questions and I am grateful for their thorough response, and I am happy to recommend publication in Nature. It is fully up to the authors but I would recommend including into the methods the discussion of the spaghetti-vs-blockade scaling properties (as well as opportunities to circumvent) discussed in their reply, as it is quite foundational analysis to the future of multi-qubit gates with Rydberg. I congratulate the authors on their excellent results.

Referee #2 (Remarks to the Author):

I am happy with the changes and grateful for the additional information the authors have provided for my benefit and education. I am also happy to not be the difficult referee for once. I have very few comments which almost certainly did not make it into the manuscript due to length restrictions, but I will list them here in any case. Anything I make no mention of you may assume is to my satisfaction. I will be referring to comments by page in the rebuttal since they are not numbered. I am recommending publication with or without these changes included.

p. 9: As per the authors my assessments as to the "losing when using entanglement" was insufficiently precise, although it does constitute to my knowledge the experimental reality up until now. It is at the very least the sceptics' baseline doubt after 20 years of claims that squeezing is practically useful. Therefore, I felt I need to entice the authors to be more upfront about this point in the main text. They were with me, but sadly not in the main text during the review. If there is space, I encourage that still. Not everyone is as deep in the technical detail as they are.

p. 10 on l. 47 comment: I was never happy with that phrasing, it being popular in the literature or not, but given there is actually a formal definition I concede the point.

p. 11 on l. 50 comment: The QFI/Cramer Rao bound discussion is always based on the underlying assumption that the estimator is unbiased. The authors cite several works about how this is a problematic assumption when the measurement has finite bandwidth. That being said, the last sentence of the comment is exactly what I was getting at.

p. 11 on l. 67 comment: I was of course getting at OAT squeezing to GHZ as the authors correctly assumed. I find it peculiar to call every state under that interaction up until arbitrarily close to GHZ "squeezed" and every state arbitrarily close *after* "(over)squeezed". There is a volume-0 parameter set where this word is not appropriate? In any case, this is terminological differences. The authors make a compelling point on the next page. I would love to see that in the manuscript and not just as the reply, however.

p. 14 on l. 387 comment: The reply in short would fit well into the conclusion. Generally the things that did not make it into the manuscript due to space constraints that still add insight could be in a brief paragraph on supplementary discussion.

Author Rebuttals to First Revision:

Reply to referees for manuscript “Multi-qubit gates and ‘Schrodinger cat’ states in an optical clock”

July 23, 2024

Response to referee 1

> `The authors have addressed all of my questions and I am grateful for their thorough response, and I am happy to recommend publication in Nature. It is fully up to the authors but I would recommend including into the methods the discussion of the spaghetti-vs-blockade scaling properties (as well as opportunities to circumvent) discussed in their reply, as it is quite foundational analysis to the future of multi-qubit gates with Rydberg. I congratulate the authors on their excellent results.`

We thank the referee again for their thorough engagement with and favorable characterization of our manuscript. We have added a new section to the Methods discussing the scaling aspects of the Rydberg blockade and spaghetti called “Ensemble size scaling for multi-qubit Rydberg gates”.

Response to referee 2

> `I am happy with the changes and grateful for the additional information the authors have provided for my benefit and education. I am also happy to not be the difficult referee for once. I have very few comments which almost certainly did not make it into the manuscript due to length restrictions, but I will list them here in any case. Anything I make no mention of you may assume is to my satisfaction. I will be referring to comments by page in the rebuttal since they are not numbered. I am recommending publication with or without these changes included.`

We thank the referee again for their thorough engagement with and favorable characterization of our manuscript. We respond to the remaining comments point-by-point below.

> `p. 9: As per the authors my assessments as to the "losing when using entanglement" was insufficiently precise, although it does constitute to my knowledge the experimental reality up until now. It is at the very least the sceptics' baseline doubt after 20 years of claims that squeezing is practically useful. Therefore, I felt I need to entice the authors to be more upfront about this point in the main text. They were with me, but sadly not in the main text during the review. If there is space, I encourage that still. Not everyone is as deep in the technical detail as they are.`

We agree with the referee that the practical utility of entangled states for metrology is of utmost importance. This is a very broad and nuanced topic which requires careful discussion. The manuscript already addresses in detail aspects of utility related to GHZ states, specifically their fragility to single-particle decoherence (as mentioned in the introduction) and the laser phase noise issue (the motivation for the cascade section), which is the focus of the work. The referee mentions aspects such as duty cycle in the original comment, and here the "experimental reality" of squeezing (though our view is that squeezing, certainly of the electromagnetic field, is one of the few protocols where a quantum enhancement has been demonstrated for practical applications, as per the references at the end of the first paragraph); both topics are very interesting, but we feel beyond the scope of this work. In order to emphasize that practical utility is an important aspect to investigate further, we had already added the following sentence to the conclusion in our first revision:

> "Comparing different entanglement strategies, ranging from spin-squeezing to GHZ-state generation, on their practical utility, accounting for trade-offs in metrological gain and robustness, is an interesting avenue for programmable clocks."

> `p. 10 on l. 47 comment: I was never happy with that phrasing, it being popular in the literature or not, but given there is actually a formal definition I concede the point.`
> `p. 11 on l. 50 comment: The QFI/Cramer Rao bound discussion is always based on the underlying assumption that the estimator is unbiased. The authors cite several works about how this is a problematic assumption when the measurement has finite bandwidth. That being said, the last sentence of the comment is exactly what I was getting at.`

We agree that the QFI/Cramér-Rao bound discussion is not fully general. We relied on it here as it is the typical paradigm for introducing the metrological utility of GHZ states (as is being done at the relevant point in the main text). Of course, we fully agree that a finite width prior is an issue; this discussion is already present in the manuscript, specifically in the cascade section, and we have left it as is.

The last sentence in our original response referred to how the different properties of different quantum states could lead to Heisenberg precision scaling. We agree that this is an extremely interesting topic, but beyond the focus of the presented results (especially at the relevant point in the text and the space limitations); this is why we chose to remove the word "thus" and avoid further discussion here (as per our original response).

> `p. 11 on l. 67 comment: I was of course getting at OAT squeezing to GHZ as the authors correctly assumed. I find it peculiar to call every state under that interaction up until arbitrarily close to GHZ "squeezed" and every state arbitrarily close *after* "(over)squeezed". There is a volume-0 parameter set where this word is not appropriate? In any case, this is terminological differences. The authors make a compelling point on the next page. I would love to see that in the manuscript and not just as the reply, however.`

We would like to clarify that spin-squeezed states are not arbitrarily close to the GHZ state under OAT, and thus the distinction is more than a terminological difference. We suspect that our initial response was not sufficiently clear and caused a misinterpretation of what we refer to as squeezed and over-squeezed. To clarify further, we reiterate a few key points here on the dynamics under OAT $e^{-i\chi t S_z^2}$ (S_z the N-atom collective spin operator projection along z) following Ref. [1]. The squeezed states are only generated for a very short duration at early times $\chi t \lesssim 1/\sqrt{N}$ where the squeezing parameter is $\xi_{\mathrm{R}}^2 < 1$. The GHZ state, on the other hand, is produced at a specific time $\chi t = \pi/2$. The interval $1/\sqrt{N} \lesssim \chi t \lesssim \pi/2$ between these two times separates the squeezed states and the GHZ state, during which a broad class of states (loosely categorized as over-squeezed) are produced; importantly these states are not squeezed, i.e. $\xi_{\mathrm{R}}^2 > 1$, and so the squeezed states are not arbitrarily close to the GHZ state. Also note that the OAT dynamics essentially reverse after the GHZ state is produced, so neither the states immediately before nor after $\chi t = \pi/2$ are squeezed. We defer to Ref. [1] (specifically Fig. 14 and the surrounding discussion in section III.B) for further details. We hope this clarifies the distinction between squeezed states and GHZ states, even in the context of OAT.

Regarding the final point made in the original response about different detection requirements/decoding protocols for different types of entangled states, we agree that it is an important aspect of quantum metrology, and one of many aspects that have to be considered when comparing different types of entangled states for metrology. Given that this manuscript solely focuses on GHZ states, we again feel such a discussion is beyond the scope of this work. As mentioned in a previous comment, we had added a sentence in the first revision stressing that this is an interesting direction for future work.

> `p. 14 on l. 387 comment: The reply in short would fit well into the conclusion. Generally the things that did not make it into the manuscript due to space constraints that still add insight could be in a brief paragraph on supplementary discussion.`

We have modified a sentence in the conclusion to better emphasize that the improved dynamic range of the cascade technique will allow access to longer dark times:

> "Employing these GHZ states for metrology, we have performed an atom-laser frequency comparison below the SQL at a short Ramsey dark time and extended the phase estimation dynamic range with a multi-ensemble GHZ-state cascade; the latter capability restores the compatibility of large GHZ states with the long dark times available for unentangled atoms when local oscillator noise dominates, as is the case for the state-of-the-art optical atomic clocks."

References

[1] Luca Pezzè, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein. Quantum metrology with nonclassical states of atomic ensembles. *Rev. Mod. Phys.*, 90:035005, Sep 2018.